

CLASSIFICATION OF INDIVIDUALS USING HANDWRITTEN NUMERAL  
WITH GEOMETRIC MORPHOMETRIC TECHNIQUES

By

NASRUL HAKIMIE BIN ISMAIL

Dissertation submitted in partial fulfillment  
of the requirement for the Master of Science  
(Forensic Science)

August 2020

## CERTIFICATE

This is to certify that the dissertation entitled **CLASSIFICATION OF INDIVIDUALS USING HANDWRITTEN NUMERAL CHARACTERS WITH GEOMETRIC MORPHOMETRICS** is the bone fide record of research done by **Nasrul Hakimie Bin Ismail** during the period of **February 2020 to August 2020** under my supervision. I have read the dissertation and that in my opinion, it confirms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation to be submitted in partial fulfilment for degree of Master of Science in Forensic Science.

Supervisor,

  
**DR. DZULKIFLEE B. ISMAIL**  
Pensyarah  
Pusat Pengajian Sains Kesihatan  
Kampus Kesihatan  
Universiti Sains Malaysia.....  
16150 Kubang Kerian, Kelantan

Dr Dzulkiflee Ismail

Lecturer,

School of Health Sciences,

Universiti Sains Malaysia (Health Campus),

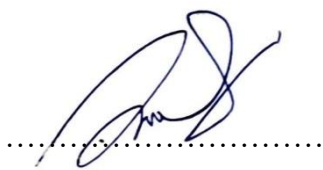
16150, Kubang Kerian,

Kelantan, Malaysia.

Date: ...9/9/2020.....

## DECLARATION

I hereby declare that this dissertation is the result of my own investigation, except where otherwise stated and duly acknowledged. I also declare that it has not been previously or concurrently submitted as whole for any other degrees at Universiti Sains Malaysia or other institutions. I grant Universiti Sains Malaysia the right to use the dissertation for teaching, research and promotional purposes.

A handwritten signature in blue ink, consisting of stylized loops and a long horizontal stroke, positioned above a dotted line.

Nasrul Hakimie Bin Ismail

Date: ...9/9/2020.....

## **ACKNOWLEDGEMENTS**

In the name of Allah SWT, the most gracious and the most merciful. All praise to Allah SWT, the Lord of universe who has given love, patience, blessing and health to complete the dissertation. Peace and blessing be upon Muhammad SAW, his family, his companions and his followers. I express my sincere thanks to my supervisor Dr Dzulkiplee Ismail and his wife as well as Co-supervisors, Dr Wan Nur Syuhaila Mat Desa and Dr Helmi Mohd Hadi Pritam. Without their guidance, persistence helps and valuable comments this dissertation cannot be completed. I also would like to thanks Kak Wan for her time to contribute knowledge in this project. Thanks to Koon, Insp Sharazi, Apiz, Fatin, Lia, Syahirah, Wanie, Wani and Zati for voluntarily contribute as subject in this project. Thank you to my parents for their prayers and financial support that give me strength to completed this project. I also would like to thanks all lecturers who taught me throughout 4 years study. Special appreciation and thanks to some of my friends for their help and support during the collection and preparation of the samples. Special thanks to all who directly or indirectly involved in performing this research. Lastly, to whoever I owe to, feel free to reach me if you need any help. Thank you.

## TABLE OF CONTENTS

CERTIFICATE .....	ii
DECLARATION .....	iii
ACKNOWLEDGEMENTS .....	iv
List of Tables.....	vii
List of Figures .....	viii
ABSTRAK .....	x
ABSTRACT .....	xii
CHAPTER 1 .....	1
INTRODUCTION .....	1
1.1 Research background .....	1
1.2 Handwriting .....	2
1.3 Numerals .....	3
1.4 Problems that affect Forensic Document Examination.....	6
1.5 Problem Statement .....	7
1.5.1 Lack of Research in Identification Using Numeral .....	7
1.5.2 New Method Approach for Numeral Recognition .....	7
1.5.3 The Need for Objectivity Approach in Research .....	8
1.5.4 The Need for Supporting Evidence .....	8
1.6 Significance of the study .....	8
1.7 Objective .....	9
1.7.1 General Objective .....	9
1.7.2 Specific Objective .....	9
LITERATURE REVIEW.....	10
2.1 Introduction .....	10
2.2 Introduction of Hindu-Arabic numerals .....	12
2.3 Introduction of Geometric Morphometric technique .....	14
2.4 Character Recognition Systems .....	20
2.4.1 Pre-processing .....	20
2.4.1.1 Binarization.....	21
2.4.1.2 Noise Removal.....	21
2.4.1.3 Baseline Detection .....	22
2.4.14 Normalisation.....	22
2.4.2 Segmentation .....	23
2.4.3 Feature Extraction .....	25
2.4.4 Classification .....	27



## **List of Tables**

Table 2.1	The differences in machine-human interactions.....	28
Table 3.1	Definition of each landmarks on the numerals.....	43
Table 4.1	Principal Component Analysis Eigenvalues Text Output for 1 .....	54
Table 4.2	Principal Component Analysis Eigenvalues Text Output for 6.....	59
Table 4.3	Principal Component Analysis Eigenvalues Text Output for 8.....	64
Table 4.4	Principal Component Analysis Eigenvalues Text Output for 9.....	69

## List of Figures

Figure 1.1	The roman number from modified symbols of V and X.....	4
Figure 1.2	The evolvement of number from Brahmi to modern.....	5
Figure 2.1	The differences between size and orientation.....	16
Figure 2.2	The visual representation of the pinocchio effect.....	19
Figure 2.3	The general steps in character recognition.....	21
Figure 3.1	Some of the components use for image enhancement.....	45
Figure 3.2	Tools of TpsUtil.....	46
Figure 3.3	Windows for setting Tps file.....	47
Figure 3.4	2 labeled menu for landmark digitization.....	48
Figure 3.5	Example of raw data of the digitize landmarks for number 8.....	48
Figure 3.6	Example of 2 panel in the outlier for number 6.....	49
Figure 3.7	Example of result for procrustes fit for number 9.....	51
Figure 3.8	Example of completed wireframe for number 6.....	52
Figure 3.9	Example of menu for PCA analysis.....	53
Figure 4.1	eigenvalues graphical result for numeral 1.....	55
Figure 4.2a	PC1 Lollipop graph for numeral 1.....	55
Figure 4.2b	PC1 wireframe graph for numeral 1.....	56
Figure 4.3a	PC2 Lollipop graph for numeral 1.....	56
Figure 4.3b	PC2 wireframe graph for numeral 1.....	56
Figure 4.4	PC Score Plot for numeral 1.....	58
Figure 4.5	Eigenvalues graphical result for numeral 6.....	60



Figure 4.6a	PC1 Lollipop graph for numeral 6.....	60
Figure 4.6b	PC1 wireframe graph for numeral 6.....	61
Figure 4.7a	PC2 Lollipop graph for numeral 6.....	61
Figure 4.7b	PC2 wireframe graph for numeral 6.....	61
Figure 4.8	PC Score Plot for numeral 6.....	63
Figure 4.9	Eigenvalues graphical result for numeral 8.....	65
Figure 4.10a	PC1 Lollipop graph for numeral 8.....	65
Figure 4.10b	PC1 wireframe graph for numeral 8.....	66
Figure 4.11a	PC2 Lollipop graph for numeral 8.....	66
Figure 4.11b	PC2 wireframe graph for numeral 8.....	67
Figure 4.12	PC Score Plot for numeral 8.....	68
Figure 4.13	Eigenvalues graphical result for numeral 9.....	70
Figure 4.14a	PC1 Lollipop graph for numeral 9.....	70
Figure 4.14b	PC1 wireframe graph for numeral 9.....	71
Figure 4.15a	PC2 Lollipop graph for numeral 9.....	71
Figure 4.15b	PC2 wireframe graph for numeral 9.....	72
Figure 4.16	PC Score Plot for numeral 9.....	73

# **INDIVIDU KLASIFIKASI BERDASARKAN ANGKA TULISAN TANGAN MENGUNAKAN TEKNIK MORFOMETRI GEOMETRI**

## **ABSTRAK**

Tiada individu yang akan mempunyai cara penulisan yang sama. Cara menulis adalah sesuatu yang akan terbentuk dan berkembang hasil dari pembelajaran yang melibatkan proses neuropsikologi dan psikomatik yang kompleks yang membentuk seseorang individu. Proses menganalisis tulisan tangan dengan tujuan untuk mengenalpasti identiti penulis telah dilakukan sejak terciptanya tulisan tangan itu sendiri. Ia telah membantu penyiasat dalam siasatan yang melibatkan analisis dokumen seperti penyelewengan, pemalsuan dokumen dan banyak lagi. Nombor yang digunakan pada masa kini adalah nombor “Hindu-Arabic” yang merupakan gabungan 10 angka iaitu 1,2,3,4,5,6,7,8,9 dan 0. Ia merupakan sistem asas sepuluh memandangkan nilai nombor meningkat pada kuasa sepuluh. Cara penulisan tangan mempunyai variasi yang tinggi dari segi saiz, bentuk dan font. Walaubagaimanapun, variasi yang paling jelas untuk setiap nombor adalah dari segi bentuk. Morfometri Geometri (GMM) adalah satu bentuk analisa melibatkan variasi di dalam bentuk. Ia menggunakan sistem koordinat kartesian geometri bukannya menggunakan pembolehubah linear, kawasan atau isipadu. Asas untuk teknik ini ialah dengan menggunakan koordinat dalam bentuk titik mercu tanda dan titik mercu tanda separa untuk merekod bentuk sesebuah gambar. Dalam projek ini, 200 sampel melibatkan nombor 1, 6,8 dan 9 telah dikumpul dari 10 orang pelajar Universiti Sains Malaysia (USM). Nombor tersebut berdasarkan hasil ujikaji sebelumnya daripada Tay Eue Kam (2019), disebabkan nombor itu berhasil untuk digunakan dalam pembahagian etnik dengan menggunakan teknik manual sepenuhnya berdasarkan kiraan serongan, lebar dan tinggi. PCA untuk PC1 dan PC2 menunjukkan peratusan varians terkumpul

untuk nombor 1, 6, 8 dan 9 adalah 100.00%, 67.59%, 56.34% and 65.17%. Hanya nombor 6 yang menunjukkan perbezaan yang tinggi dalam membezakan individual. Morfometri Geometri mempunyai potensi dalam individual identifikasi berdasarkan bentuk nombor.

# **CLASSIFICATION OF INDIVIDUALS USING HANDWRITTEN NUMERAL WITH GEOMETRIC MORPHOMETRIC TECHNIQUES**

## **ABSTRACT**

No individual will have exactly same handwriting. Handwriting style is something that will acquire and develop through years of learning involving questioned that form individuality. The analysis of handwriting for the purposes of author identification had started since the origin of the handwriting itself. It assisted investigator in crime involving document analysis such as fraud, document falsification and many more. The numbers that used nowadays are also called Hindu-Arabic numbers are combination of 10 digits which are 1,2,3,4,5,,6,7,8,9,and 0. It is a base ten system since the value increases by the power of ten. The handwriting style has very high variations sizes, shapes and fonts. However, the most distinctive characteristics is the shape itself. Geometric morphometric (GMM) is a study of variations in shape. It uses Cartesian geometric coordinates rather than linear, area or volumetric variables. The basis of the technique is coordinates of identifiable landmarks (LM) and semilandmarks (SL) to capture the image shape. In this project, 200 samples of consisting of numeral 1, 6, 8 and 9 were collected from 10 students of Universiti Sains Malaysia (USM). The number is based on previous study by Tay Eue Kam (2019), as it successfully used in ethnic discrimination by using fully manual method based on slant, width and height measurement. The PCA of PC1 and PC2 shows cumulative variance percentage for numeral 1, 6, 8 and 9 were 100.00%, 67.59%, 56.34% and 65.17% respectively. Numeral 6 was the only numerals that show significant differences between individuals. Geometric morphometric is useful in individual identification based on shape of the numerals.

## **CHAPTER 1**

### **INTRODUCTION**

#### **1.1 Research background**

No individual will have exactly same handwriting (Tarannum et al., 2015). The analysis of handwriting for the purposes of author identification had start since the origin of the handwriting itself. It assist investigator in crime involving questioned document such as fraud, unknown letter and many more (Sargur et al., 2003). Forensic document examiner is an area that is well develops in Forensic Science field. Question document examination involves a comparison of the document, aspect of the document with the known standards (Songer, 2015). The standard procedures for forensic document analysis are questioned samples, collected standards and reference standards (Moszczynski, 2019). Collected standard can be define as handwriting samples that been written in normal course while the reference standard can be define as handwriting samples that been written when requested by the investigator or the examiner (Gilmour and Bradford, 1987). Forensic document examiner have establish numerous techniques to answer set of questions such as using a known or a set of known document to link with questioned document, identify either the handwriting is forged and whether the handwriting is disguised (Sarguret al., 2003). The most common aspect that related to forensic document analysis are handwriting and signature analysis. The documents that contain handwritten notes with written words, numerals and signatures were examined for writer's identifications (Hayes, 2006).

## **1.2 Handwriting**

Handwritings can be defined as any record or presence of letters of the alphabet, punctuation, spaces to convey any thought or ideas that usually write using a pen or pencil for handwriting and keyboard for typing (Englishclub, 2020). People deliver message to the others either in permanent or semi-permanent mode through penmanship system which is writing (Koppenhaver, 2007).

Handwriting style is something that will develop and acquired through years of learning involving complex neuropsychological and psychosomatic process that form individuality. People start to write in school simply by copying graphical signs at first but it will slowly diminish and develop their own writing style which is influence by the individual and the habit (Moszczynski, 2019).

This is because as people grown up, graphic maturity is reached when the individual motor skill is fully develop and the individual can eventually just focus on the content, not the act of the writing. The writer can concentrate more on what they write and let his or her subconscious mind to handle the act of writing as it become habit and the method of construction of various letters combinations an words is set (Koppenhaver, 2007).

There are several factors that can affect how people write such as hand-eye coordination, wrist movement flexibility, the way individual grip the writing instruments as well as attitude and discipline (Koppenhaver, 2007). As the people grown up handwriting will change throughout the lifetime until it achieve maturity, stable and show senility and it will remain as the individual characteristic of the person (Ta et al., 2010). However, in some case there are still probability for handwriting to change due to certain aspects either temporary or permanent. Factors

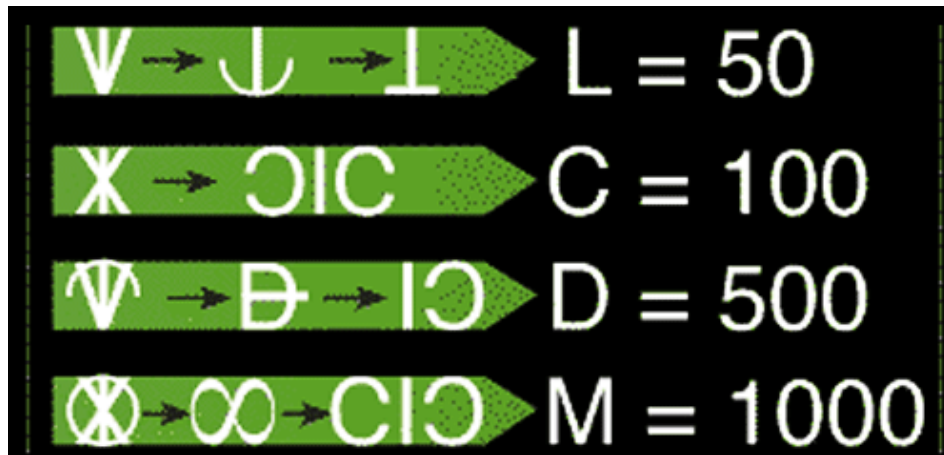
such as uncomfortable writing position, uncommon writing instruments, use of certain style, physical and emotional condition, intoxication by alcohol, tiredness and mental illness such as schizophrenia can influence how people write (Moszczynski, 2019).

Handwriting can be characterise into class and individual characteristics. Class characteristics are characteristics that can be found in the group of people due to penmanship system that the writer learned while individual characteristics are deviation from the penmanship system as they modify and stylize their writing. The characteristics such as direction, slant, rhythm, pressure, line quality, speed, size, proportions and spacing, size, proportions of the letter and baseline alignment are unique that can give arise to individual characteristics beneficial for person identification (Koppenhaver, 2007).

### **1.3 Numerals**

People had use various of techniques to count and describe amount since early of civilization by using notches in a tree, sticks, stones and knot tied into ropes to indicate amounts (Koppenhaver, 2007). Back on the historic era, over past of 5,500 years, human had create more than 100 images, relatively permanent and largely non-linguistic means for representing numbers (Chrisomalis, 2018).

Before the numbering system that we used nowadays, people used Roman Figure which was Etruscan origin that based on biquinary system until XVI century (Sarcone and waeber, 2020). The letters 1 to 3 based on the lines, letter 4 is actually 5-1, 'V' represent as 5, 10 is actually two letter 'V' where one letter was inverted that become 'X' same with L represent 50 C represent 100, D represent 500 and M represent 1000.



**Figure 1.1:** the roman number from modified symbols of V and X (Sarcone and Waeber, 2020)

The numbers that we used nowadays are also called Hindu-Arabic numbers are combination of 10 digits which are 1,2,3,4,5,,6,7,8,9,and 0. It is a base ten system since the value increases by the power of ten (Sarcone and waeber, 2020). These digits were introduced from India to Europe within XII century by Leonardo Pisano, an Italian mathematician (Sarcone and Waeber, 2020).

During third century Before Common Era (BCE), Leonardo had visited India and change the number system that been used currently at that time which was Brahmi. The Brahmi numerals were complicated hence changes of shapes were made into Gupta numerals. It was created during Gupta dynasty in fourth century. However, the Gupta numerals have issue which is take a long time to write. Hence, the numerals were change from Gupta to Nagari numerals which was in cursive forms. The Figure 1.2 shows that the Hindu row is an example of Nagari numerals. In eighth century, during Islamic invasion in Northern part of India, these numbers were taken by Arab and further evolve and spread to the whole world (O' Connor and Robertson, 2000).



Brahmi	↓		—	=	≡	+	୮	୧	୨	୩	୪
Hindu	↓	୦	୧	୨	୩	୪	୫	୬	୭	୮	୯
Arabic	↓	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
Medieval	↓	୦	୧	୨	୩	୪	୫	୬	୭	୮	୯
Modern		0	1	2	3	4	5	6	7	8	9

© G. Sarcone, [www.archimedes-lab.org](http://www.archimedes-lab.org)

**Figure 1.2:** The evolvement of number from Brahmi to modern (Sarcone and Waeber, 2020)

From the Figure 1.2, zero was start in Hindu and then Arab translated the word sunya into their own word sifir which still in the same meaning and then Latinized by Italy to zephirum in the thirteenth century. Various changes occur until it finalized with word zero (Buddhue, 1941).

The styles of writing numbers have very high variations sizes, shapes and fonts. For numeral, specialised field of research named optical character recognition (OCR) is develops to identify handwritten numbers in automated or semi-automated manner (Hanmandlu and Murthy, 2007).

Handwriting cases involving numerals commonly can be found in cases that related to financial such as alteration in checks. Example case that previously happen was alteration of cash \$1,000 by changing the numeral “1” with “9” and adding the recipient name (Forensicsciencesimplified, 2013).

Although people tend to change their handwriting for disguise purpose, most people does not change the way their write numbers. Hence, many cases can be solve

through comparison due to vivid differences between number such as people make zeros smaller than other numbers (Koppenhaver, 2007).

#### **1.4 Problems that affect Forensic Document Examination**

The challenges that been face by the forensic document examiner is multi-individuality of handwriting where one person can write two distinctive different handwriting styles. This is problematic as it can lead to the handwritten is from different author (Moszczynski, 2019). The other problems is disguised handwriting in order to make handwriting have less characteristics of the individual. Sometimes disguised handwriting can easily be recognized due to stiffness and artificial look (Konstantinidis, 1987).

Moreover, if the numbers of potential writers are too large it will be very hard to match and compare between question and reference documents (Bensefia et al., 2016). Despite well established and long history of development of forensic document examiner, the analysis is still difficult, time consuming and subjective as qualified expert evaluate the analysis such as letters, strokes and writing styles based on their experience. Other problems are inevitable variations due to psychological problem, emotional state, moods, illness, drug and medications, poor handwriting quality large time gap between the questioned or incriminated samples with reference samples can affect the judgement of the expert (Gluhchev, 2004).

Mechanical factors such as defect in writing instruments, the quality of the paper and the writing surfaces, lighting, the position of the writer can affect the handwriting (Koppenhaver, 2007). In the case of check forgery addition to original number is the most common problems for example one can be made into four, seven

or nine, two can be turn into three or five and zeros can be added to increase the amount to hundreds or thousands (Koppenhaver, 2007).

## **1.5 Problem Statement**

This study is designs to identify the capability of geometric morphometric approaches in accuracy and precision for discrimination between individuals using their handwritten numerals. If this technique is successfully identify and discriminate the individuals, other scope such as from set of number, which number represent significant difference that can be used to discriminate the individuals can be identify.

### **1.5.1 Lack of Research in Identification Using Numeral**

The studies for identification of the author using the alphabets and signature are quite common in forensic document examination. The study related to numerals is quite few and it can be numeral itself can be differentiates into multiple scope. Example of study in numeral is Recognition of Handwritten Bangla Numerals using Adaptive Coefficient Matching Technique by Amitava Choudhury (Choudhury et al., 2016). In contrast this study will focus on the classification of the author using handwritten Arabic numeral.

### **1.5.2 New Method Approach for Numeral Recognition**

There are several journal related to classification of the individuals using numeral. However different methods approach were used for the author identification. For example the study of recognition of handwritten numerals focused on using Fuzzy model method (Hanmandlu and Murthy, 2007), study of recognition of handwritten numeral using Gabor features (Hamamoto et al., 1996) and numeral recognition using PCA mixture model (Kim et al., 2002). In contrast, this study will

focus in using geometric and morphometric techniques for the classification of the individuals.

### **1.5.3 The Need for Objectivity Approach in Research**

The most important things during presenting the evidence in the court is to convince the judge and the lawyer about the importance of the evidence. In case related to question document examination, the investigation itself is more to qualitative and forensic document examiner have to show how certainty of them with the conclusion to avoid immutability of the conclusion. Before the examination is concluded with verbal scale either 'inconclusive' or 'consistent with' the information in the numerical representation which is quantitative representation can be shown to increase the evidential value of the evidence (Neumann *et al.*, 2016).

### **1.5.4 The Need for Supporting Evidence**

This study is important as it can assist in the investigation in interpretation of the observation in cases related to question documents as it can add value and link to the other evidences (Evetts *et al.*, 2000).

## **1.6 Significance of the study**

The purpose of the study is to answer the question if it possible to use geometric and morphometric techniques for classification of the individuals using handwritten numeral characters. The scope of the study is focus on the number 0,6,8 and 9 and the sample taken from subject in Universiti Sains Malaysia (USM), Kubang Kerian.

The study is very important in the field of forensic handwriting examination. It will be very beneficial in cases related for identification of the writer using

numeral characters. Moreover geometric morphometric techniques are techniques that can be mastered by everyone and give different approach in solving cases related to question documents.

## **1.7 Objective**

### **1.7.1 General Objective**

Discriminate the individuals by their handwritten numeral characters using geometric morphometric technique.

### **1.7.2 Specific Objective**

#### **Objective 1**

To analyses the handwritten numeral 1, 6, 8 and 9 for different individual characterization.

#### **Objective 2**

To apply geometric morphometric techniques approach for individual characterization.

#### **Objective 3**

To identify which numeral from numeral 1, 6, 8 and 9 that have potential for discrimination between individuals.

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 Introduction**

Question document examination is one the branch in Forensic Science study (Swgdoc, 2018). Since the late nineteenth century, researchers have searched for significant methods and analysis for individual identifications in this area (Bird, Found, Ballantyne, and Rogers, 2010). During the past 30 years, the value of scientific knowledge in this field has become more crucial (Risinger, Denbeaux, and Saks, 1989; Saran, Kumar, Gupta, and Ahmad, 2013; Mnookin, 2001). It commonly focused on the authenticity or doubtful presence in documents such as handwriting signature, typewriting, printing process and other marks that deviate from originality of documents Forensic document examiner deal with the authenticity of the question documents in crime cases such as forgeries, counterfeiting, cheque fraud, identity fraud, suicide notes, extortion, contested wills and dead lawsuits and many more (Swgdoc, 2018).

Identity fraud document and security documents are commonly used in crime such as terrorism activities, trafficking of the immigrants, scams such as cheque fraud, facilitate smuggling of things such as drugs, weapons and cross borders crime. There are numerous cases every year dealing with evidence such as wills and ransom notes. Forensic document examiner have important role in immigration, border control security and forensic document examination facility as they possessed knowledge in examination and comparison of handwriting, typewriting, printing, ink analysis and other documents (Rudner, 2008; Hassaïne et al., 2012).

No individual will have exactly the same handwriting as it is unique (Kedar et al., 2015). Since the late nineteenth century, researchers have searched for significant method and analysis for individual identification in this area (Bird, Found, Ballantyne, and Rogers, 2010). The idea of uniqueness started in the nineteenth century with Quetelet who hypothesizes “nature never repeats” that give arise to product rule probability that suggest great odds against such repetition (Cole, 2009).

There were numerous studies that had been done in order to carry out features and characteristics extraction for handwriting classification, authentication and identification. Compared to the electronic and printed document, handwriting itself gives additional information for example in determination of the personality of the writer. From handwriting itself, individual characteristics and features form signature, numeral and alphabets can be extracted to represent the author (Siddiqi and Vincent, 2007). This is due to presence of natural variations that caused deviation between handwriting characteristics (Koppenhaver, 2007). The variations is handwriting styles that develop and acquired through years of learning involving complex neuropsychological and psychosomatic process that form individuality (Moszczynski, 2018). In forensic science application, this types of studies are important for process distinction of the individuals (Siddiqi and Vincent, 2007). This fact then is exploit by the forensic document examiners that provided legal evidence authenticity into the court of law (Garz et al., 2016).

There are multiple studies that show the importance of individuality of handwriting. The main objective of analysis of forensic identification is individualization, placing an object into single, solitary unit (Saks and Faigman, 2008). For example, Srihari et al. (2003), stated that it is possible to individualize the writer by observing the degree of variability in the handwritten numerals. Other

studies from Alford (1965), to strengthen the concept of individuality in the handwriting, although the person tried to mask or attempted disguised handwriting, it will be burden mentally hence they may overlook or neglect in the formation of numerals. This shows that identification through comparison is important in forensic document examination. Although people tend to change their handwriting to disguise, most people not change the ways their write numbers. Hence, many cases can be solved through comparison due to vivid differences between numbers such as people make zeros smaller than other numbers (Koppenhaver, 2007).

This study applied geometric morphometrics hence is focusing on form which are size and shape. Size indicates the scale of the objects either it is big or small. Once the size and the position is determined, shape is analyzed. As world growth, the technology become more develop and sophisticated, the crime such as fraud become more creative and varied, new things and changes happen around us. Hence, research and new analysis need to be done in order to keep updated, improve and improvise so would not left behind, move coherently with the current development.

## **2.2 Introduction of Hindu-Arabic numerals**

For the history of number, number had evolved over time and is influenced by the era and culture. From the world history the history of number was discovered and rediscovered four times. Firstly, was in the second millennium before the Common Era (BCE) in Babylon based on the essential tools of mathematics. Secondly was in the Common Era rediscovered by mathematic arithmetician. Thirdly was in the fifth centuries of the Common Era by the Mayan astronomers and lastly rediscovered by the India (Ifrah, 2000).



There are multiple numbering system such as Bangla, Devanagari, Roman, Urdu and Hindu-Arabic (Obaidullah et al., 2015). However, the most common numeral system that is being used worldwide is Hindu-Arabic numeral system which consists of ten symbols which are 0,1,2,3,4,5,6,7,8 and 9 (Danna, 2019). It is a base ten system since the value increases by the power of ten (Sarcone and waeber, 2020). Hindu-Arabic numeral was first developed in the East around the 5th century and brought to the Europe across the Mediterranean during late medieval period (Kunitzsch, 2003; Chrisomalis 2010). In 10 to 13th centuries, Europe was introduced to the Hindu-Arabic numerals from India (Høyrup, 2012). Although during that era, Europe had used roman numbers, the reason they shifted to Hindu-Arabic numerals due to usefulness and easy to be apply in multiplication and division especially in the international trade when compared to Roman. This is because calculation in multiplication and division are important in exchange and conversion rates (Otis, 2017). The application of arithmetic in Europe from late 13th century to 1600 can be proven through original database recording from 1280 texts written by 340 authors (Danna, 2019).

In 1202, a merchant mathematician, Leonardo Pisano or known as Fibonacci, had completed his studies and named it *Liber abaci* (written in Latin) as an effort to summarized and spread Hindu-Arabic mathematic to the European. It content of functioning of the positioning numeral system, summary of Hindu-Arabic mathematics at that time and how mathematics can be applied in order to solve practical and commercial problems. The practical component of the document was the root in development of mathematic application nowadays (Danna, 2019).

### **2.3 Introduction of Geometric Morphometric technique**

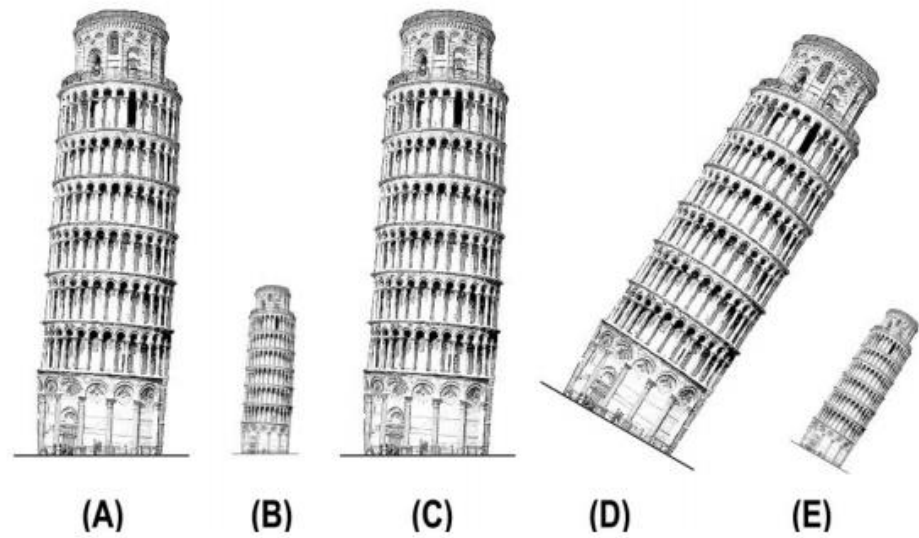
Geometric morphometric (GMM) is a study of variations in shape and covariations between variables (Adams, Rohlf and Slice, 2004). This study started at the end of 19th and the beginning of 20th century in the study of phenotypic trait. The standard analytical techniques for morphometric were developed by the English school of biometrician lead by Galton and Pearson. They developed techniques such as correlation coefficient (Pearson, 1895), linear regression and principal component analysis (Pearson, 1901; Hotelling, 1933). Fisher and Mahalanobis then assisted them in the development of modern discipline of statistics such as discriminant analysis and variance (Fisher, 1935).

Traditional Morphometric (TMM) was fully used in the second half of the 20th century (Marcus, 1990). It used was based on the application of standard multivariate analysis of arbitrary collections of distance measures, ratios, and angles (Rohlf, 2002). However the TMM need to be improvised due to three main weaknesses. Firstly, it was very hard to separate information between size and shape for example in allometry. To solve this problem, ratio cannot be used as different shape can have the same ratio. Secondly, the spatial relationship cannot be preserved although it use measurement between points. Lastly the result was in the form of Table of measurements or coefficients make it interpretation of the results is difficult due to not easily related to the original morphologies. Problems arises lead to change called geometric morphometric revolution (Rohlf and Marcus, 1993). In 1917, Thompson had proposed that the use of grid can used to compare shape in biological structures on attempt to improvise. In 1980, scientists had successfully developed technique to measure and visualize difference which is superimposition or specifically prosecutes superimposition (Cardini, 2013).

The GMM technique is a statistical technique that can be applied in two and three spatial dimensions (Fruciano, 2016). The degree of covariations between parts of the structure usually will be studied using geometric morphometric method (Klingenberg, 2009).

The basis of this technique is coordinates of identifiable landmarks (LM) and semilandmarks (SL). The landmarks created are not equal (Slice, 2007). The difference between LM and SL is that the SL is used as labelling in smooth curves or surface that cannot be clearly identified (Lorenz et al., 2017). LM also can be called as a set of homologous points while SL is not strictly homologous but retain its positional correspondence. In this analysis, data acquired either will be in form of x, y, z coordinate for 3 dimensional or x, y for 2 dimensional. The data then usually will be analysed using general linear models or principal component analysis (Fruciano, 2016)

Shape can be defined as the geometric features of the object without taking into account position, size and their orientations (Dryden and Mardia, 2016). Hence, the shape for example numerals can be observed without worrying about their position and size for visualization in the morphometric analysis (Rudemo, 2000). Figure 2.1 shows the example of differences of position and size for better understanding. (A) is the original image of Pisa tower. (B) is the translation and size differences, (C) is translation, (D) is translation and the object is rotated and lastly (D) is translation, size changes and rotation. From the landmarks position, to extract the information from the shape by removing extra information such as orientation, size and position the process called Procrustes superimposition is applied (Klingenberg, 2010).



**Figure 2.1:** The differences between size and orientation.

Due to prosecutes superimposition process that been applied, the coordination of the landmark on each shape probably will be varied and show variations in shape. The variations of shape for any given number of landmarks can be characterized into shape spaces. It is a special type of morphospace where each points of the LM or SL represents the shape and the distance between each point related to changes between the shapes. To observe the processed the variations shape, multivariate analysis need to be done as it is a method that can simultaneously analyse the covariations of all landmark coordinates (Klingenberg, 2010).

Common digitizer software that been used specifically for landmark coordination is tpsDIG2, by tps software designed by Rohlf or geomorph package. Other method such as image processing software, ImageJ also can be used for landmark coordination of the images. For tpsDIG2, it has slight advantage over ImageJ as the image with digitize landmark will be directly converted into tps format

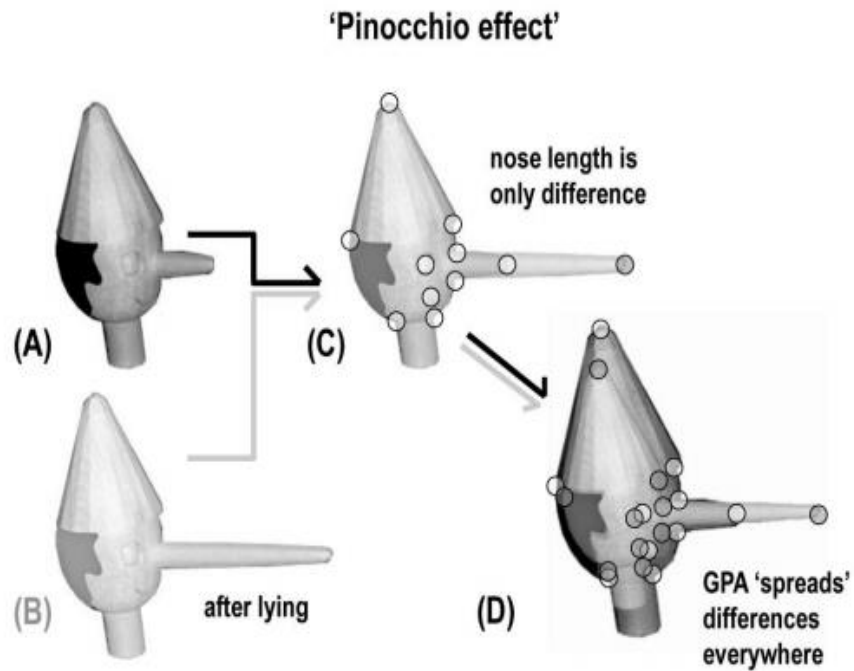
for further analysis of geometric morphometric (Tatsuta, Takahashi and Sakamaki, 2018).

There are multiple multivariate analysis that can be used such as Principal Component Analysis (PCA) for the purpose of analysis the distribution and the pattern in data variations in order to show correlation between data. Next, canonical variate analysis can be used for separation between groups by observing new shape variables to show variations within groups. Multivariate regression will be used if the analysis want to observe allometry or evolutionary of shape changes over time (Klingenberg, 2010). The concept of this analysis is measuring the degree of independent variable (predictors) and dependent variable (responses) nearly related. Lastly, Partial least square analysis will be used to show covariations between shapes. The analysis focus on finding the optimal variables to shows patterns of covariations. The most distinctive variables that show covariations between two set of variables will be identified for example in study of shape between anatomical structures (Klingenberg, 2010).

One of the main differences using geometric morphometric technique is the result can be visualized and depicts shape changes based on illustration or computer animations (Rohlf and Marcus, 1993). Shape changes are the difference that presence between shapes for example sex dimorphism (shape change from female to male or male to female). There were multiple studies for different kind of visualization for shape and changes in shape especially in scope of evolution and development of organisms used in geometric morphometric. Mainly, visualization of shape changes follow two principles. The principles either visualization of shape change by looking at the relative displacements of the corresponding landmarks in multiple shapes or by observing the deformation of a regular grid, an outline or a

surface interpolate from the changes of the shape. To analyse the shape changes, the interpretation of the results is based on relative displacements of the landmarks in the whole structure of the shapes (Klingenberg, 2013).

There is situation or error that can happen where extremely localized deformation of the superimposition. The variations is significantly localized to a small region or group of landmark or a single landmark. This is widely known as “Pinocchio effect”. This commonly happen because of least square procrustes superimposition has tendency to distribute the variations of landmarks from higher variations to the lower variations (Klingenberg, 2013). Based on the Figure 2.2 (A) is before and (B) is after lying. The differences that can be observed is on the nose represent b the grey landmark in (C). However, after proscrutes superimposition, the differences will spread all over the head (Cardini, 2013). Implication of this error is not very serious in study for shape quantification related to analysis of morphological differences between groups because geometric morphometric technique still can give accurate data and will discard the effect. However in study such as morphogenesis, it focuses on localization of the shape deformation (Tatsuta, Takahashi and Sakamaki, 2017).



**Figure 2.2:** The visual representation of the pinocchio effect.

MorphoJ is the most common software that uses for superimposed and basic geometric morphometric analysis. This is because the software provided with user friendly interface that does not require programming skills as well as wide range of analysis. In previous study of phenotypic and genotypic testing by Baranov (2018), MorphoJ was used to study the features of asymmetry of the shape of the laminae of English oak (*Quercus robur*). Fluctuating asymmetry (FA) and direct asymmetry (DA) were analysed in order to identify characteristics of the population variability. Landmarks were placed on the structures and the FA value of the shape of the tree was estimated based on the positioning of the landmarks in the Cartesian coordinate. This approach gets positive feedback as it was suitable in studies at different level of biosystems and biomedical aspects (Baranov, 2018). For superimposed, other software that can be used are tpsSuper, geomorph, shapes, CoordGen and SuperPoser. For basic geometric morphometric MorphoJ is specializes in phylogenetic, quantitative genetics and analyze modularity of shape in data. Other

software such as Evomorph by Cabrera and Giri specializes in evolutionary process simulation using geometric and morphometric data and TNT by Golobloff for calculation of most parsimonious cladogram based on landmark coordination (Tatsuta, Takahashi and Sakamaki, 2018).

## **2.4 Character Recognition Systems**

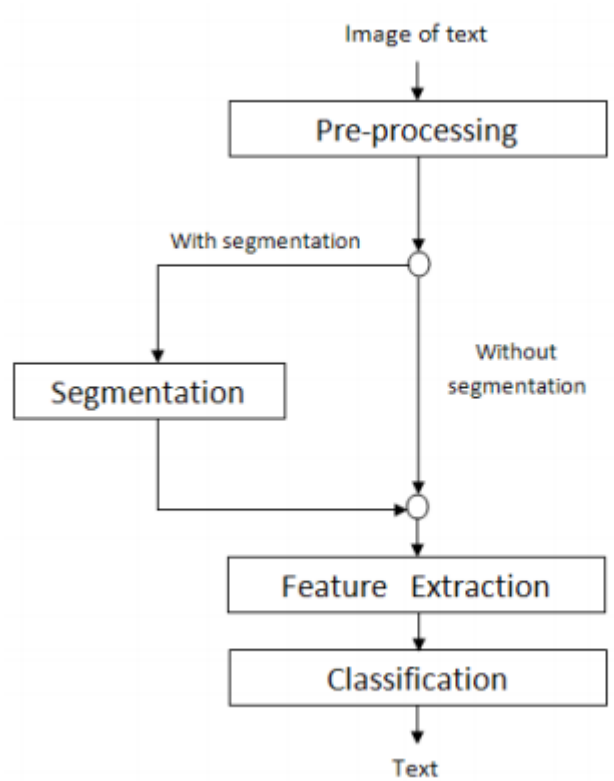
In this era of technologies, handwriting identification or characters recognition can be divided into two parts either offline or online categories. For online categories, it is dynamic method where writing samples are taken from electronic gadgets such as Tablets, PDA's, magnetic pad, smartphones resulting sequential and spatial information. Offline writing identification is a static method where writing samples are taken from scanned documents hence less sequential information (Rehman et al., 2018). Plamondon and Srihari (2000), had specifically conducted comprehensive survey in differences of offline and online handwriting. The offline writing identification then can further be divided into two categories, printed character recognition and handwritten character recognition. Printed characters have same styles and sizes while handwritten characters styles and sizes can be varied based on the writer (Lawgali, 2015).

### **2.4.1 Pre-processing**

Basically there are four general steps for handwritten character recognition starting from pre-processing, segmentation, feature extraction and classification (refer Figure 2.3). The pre-processing is important and initial step in any recognition system (El Abed and Märgner, 2007). The purpose of this step is to enhance the desire image of the text by removing any background or noise that can affect the recognition process (Farooq, Venu Govindaraju and Perrone, 2005). There are



several tasks involved such as binarisation, noise removal, baseline detection and normalisation (Lawgali, 2015).



**Figure 2.3:** the general steps in character recognition

#### **2.4.1.1 Binarization**

Binarisation is a process where the image is convert into binary format. Values for the background are pixel as 1 (white) while for foreground pixel as 0 (black). One of the advantages of this technique is increase the processing speed (Lawgali, 2015).

#### **2.4.1.2 Noise Removal**

When images are scan, distortion at the background often presence as noise. It is unwanted part of the writing that can interfere with the recognition process. There are several ways for noise removal such as filtering and morphological operation (Arica and Yarman-Vural, 2001). The extension of filtering is smoothing which is

mathematical morphology operation to reduce noise. Smoothing work in two ways, either separating interconnected objects or fill any empty spaces. Both of the ways apply morphology operations of erosion and dilation respectively. The morphological operation can work in multiple ways such as smoothing the contours, connect the strokes and extract the boundaries (Lawgali, 2015).

#### **2.4.1.3 Baseline Detection**

Baseline can be defined as an imaginary horizontal line that connects the character of words (Lorigo and Govindaraju, 2006). Baseline is important feature for slant correction and characters segmentation (AL-Shatnawi and Omar, 2009). There are multiple methods use by researcher for baseline detection. El-Hajj, Likforman-Sulem and Mokbel (2005) used horizontal projection method for detecting baseline. Pechwitz and Margner (2002) used method based on the word skeleton for baseline detection. Farooq et al. (2005), improvised by using word contour instead of word skeleton for baseline detection. Burrow (2004), proposed more complicated approach of detecting baseline based on distribution of binary 1 and 0 by using Principal Component Analysis (PCA).

#### **2.4.1.4 Normalisation**

Normalisation is the process of standardisation due to the high variations in handwriting such as styles and sizes of the characters. Size normalisation is use to reduce the variations in size and to adjust the size of the character (Liu, Cai and Buse, 2003). Example approach that had been done is dividing the character into zones then scales each zone separately. During scanning, one of the possibilities can happen on the image is slightly tilt or rotate. This problem can be detects by observation on the baseline. Skew correction is use to correct the angle orientation

(Lawgali, 2015). Other method proposed by Jian-xiong Dong et al. (2005), they used algorithm to correct skew and slant based on Radon transform.

### **2.4.2 Segmentation**

Segmentation is a process where the text are segmented into smaller part or unit such lines, characters and words. The segmentation process needs to be done as it affects recognition rate (Alginahi, 2013). There are 3 types of segmentations that can be performed segmenting a page into lines, segmenting a line into word and segmenting a word into characters.

- Segmenting a page into line: Paragraph consist of multiple lines. This lines can be separated to minimize the text forms. One of way to achieve this is through horizontal projection (Aghbari and Brook, 2009).
- Segmenting a line into word: After segmenting into line, it can be further minimized into words. This part can be categorized into 2 categories word and sub-word depends on the space between the word. Line with many spaces will be separated into word while lesser space into sub-word (Kim, Govindaraju and Srihari, 1999).
- Segmenting a word into character: This is the smallest form can be achieve where word minimize into single character. Character can be varied in many forms. Therefore, segmentation points standardized at the end of character and the beginning of the next character (Lorigo and Govindaraju, 2005).

There are several techniques for character segmentation that commonly used (Alginahi, 2013).

- Vertical projection: The vertical projection converts two-dimensional information into one-dimension. This technique is based on the concept that the line interconnect between character is thinner compare to other parts. Some of the researcher used algorithm in this technique for character segmentation. The differences between vertical and horizontal projection is it commonly used for words, sub-words and character segmentation while horizontal projection for line segmentation and baseline extraction.
- Thinning: The information in shape of the characters can be obtained from the skeleton of the characters. There are researches that used algorithm for skeletons extraction. For example, Cowel and Hussain (2001) use thinning algorithm with post processing to produce thinned form of characters. They combined with other techniques to produce better segmentation results as way to solve issue of thinning in poor quality images.
- Contour tracing: This technique is applied by following the outer part, contour of the characters. Romeo-Pakker *et al.* (1995), tried to overcome overlapping upper or lower handwritten character by applying contour tracing algorithm. They divided the process into 2 stages, detected the connection strokes between the characters based on differences in the line thickness. Next, upper contour detection for primary segmentation points identifications.
- Artificial Neural Network (ANN): ANN is used to verify the valid segmentation points. Hamid and Haraty (2001).identified pre-segmentation by using topographic features of connected block of characters. Then ANN was used to verify for valid and invalid segmentation points.